DOCUMENT RESUME

ED 207 256 EA 014 124

AUTHOR	Louis, Karen Seashore
TITLE	Policy Researcher as Sleuth: New Approaches to Integrating Qualitative and Quantitative Methods.
INSTITUTION	Abt Associates, Inc., Cambridge, Mass.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C. Dissemination and Improvement of Practice Program.
PUB DATE	16 Apr 81
CONTRACT	400-78-0002
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981). For related documents, see ED 188 314, ED 192 437, ED 196 134, EA 014 118-131 and EA 014 167.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Case Studies; *Data Analysis; Databases; Data Collection; Demonstration Programs; Elementary Secondary Education; Federal Programs; Policy Formation; *Program Evaluation; Program Implementation; Public Policy; *Research Design; Researchers; *Research Methodology; Research Problems; Research Utilization; *Social Science Research
IDENTIFIERS	Consolidated Coding; *Interactive Model; *Research and Development Utilization Program

ABSTRACT

The movement toward integration of qualitative and quantitative research methods within the same study has been most evident in social policy research. Four major approaches to integrating can be identified in recent studies. The sequential, the parallel, and the fused models have been rather widely used. The interactive model, developed during the course of the study of the Research and Development Utilization (RDU) program is relatively new. The first three models are presented briefly, and the interactive model is discussed in greater depth, with attention given to how it evolved during the course of the RDU study. The use of "consolidated coding" to merge data collected by survey with that collected through semi-structured site visits is discussed. Some of the issues and problems associated with the research methods are presented. (Author/MLF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED207256

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

POLICY RESEARCHER AS SLEUTH: NEW APPROACHES TO INTEGRATING QUALITATIVE AND QUANTITATIVE METHODS

Karen Seashore Louis

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

EA 014 121

April, 1981

Paper presented at the annual meetings of the American Educational Research Association on April 16, 1980 in Los Angeles, California.

This paper was prepared with support from Contract 400-78-0002 with the National Institute of Education. While NIE encourages dissemination of findings from research funded by the agency, this report does not necessarily represent the opinions of the agency.

POLICY RESEARCHER AS SLEUTH:

NEW APPROCHES TO INTEGRATING QUALITATIVE AND QUANTITATIVE METHODS

For the past thirty years, it has been common to refer to the existance of two distinctive "paradigms" governing the methods of social science enquiry. The first paradigm stresses the need to apply research design and analysis principles derived from the "hard" sciences, and emphasizes the desireability of experimental or quasi-experimental design, and statistical analysis. A second paradigm argues that social phenomena are essentially different from those observed by the hard sciences and that, in order to understand them adequately, we must understand the ways in which they occur naturally, and their meaning to members of the social structure. A "holistic" understanding of human social structures and behaviors requires a qualitative, observationally based methodology rather than experimental manipulation and analysis of a selected number of variables. As recently as 1977, one observer of these two camps commented that the gulf between them was so great that it was unrealistic to assume that there would be any "grand synthesis" in the forseeable future, and that any steps toward synthesis were on the "fringes" of paradigms.

However, there are a number of indications that a felt need for something more than simple detente betweer the camps was growing. Some experimental methodologists, for example, have recently taken tentative steps toward not only acknowledging the existance of an alternative paradigm, but also its suitability for studying phenomenon which have typically been dominated by quantitative approaches (Campbell, 1974; Cook and Cook, 1977). Similarly, researchers who are advocates of qualitative methods have also called for greater attention to standardization of analysis procedures (Sieber, n.d.). Finally, a number of key articles and books have advocated for integration between qualitative and quantitative methods within the same study (Sieber, 1975; Lazarsfeld, 1976; Cook and Reichardt, 1979). The expressed need for integration is not occurring only at the fringe of social sciences disciplines, but is also supported at the center, and is becoming more widespread. The tidiness of trhe divisions between camps is clearly

breaking down: one can no longer assume that an Anthropologist is totally ignorant of statistics, and traditional experimental psychologists are hotly discussing the problems of small N designs for which preferred statistical techniques are in appropriate (Herson and Barlow, 1976; Krachtowill, 1977).

The movement toward integration between qualitative and quantitative methods has been fostered most evidently in social policy research for several reasons. First, some of the early high aspirations for quantitative social policy research were deflated by an accumulation of "null" findings, and "black box" research designs were unable to reveal why apparently massive experimental treatments should produce no measureable effects. Thus, committed empiricists began to look at qualitative research methods as an approach that might help them to improve their analysis -- to point to interaction effects that should be explored, to allow them to account for otherwise inexplicable findings, and so forth.* Other researchers, such as Gross, et al., (1972) or Charters and Pellegrin (1972), raised the need to examine whether or not a treatment had actually been implemented as part of an elaborate experimental or quasi-experimental design. These studies also indicated the difficulty of determining the degree of implementation without some qualitative understanding of what constituted implementation for a given program.

Second, the most rigorous and sophisticated of designs has not eliminated doubts about the durability of policy research findings. Rather than eliminating controversy over the results of social policy and evaluation research, rigorous designs have simply raised a new dimension to the debate. Any observer who does not like the results of a major policy study can almost invariably find a variety of methodological or analytic flaws which undermine its validity. Not surprisingly, some policy makers have arrived at a deep seated skepticism about the durability of supposedly "hard" findings -- at

*See Sieber (1975) for an extensive discussion of the ways in which qualitative research can be used to complement a design which is predominantly quantitative.

least where they are unsupported by qualitative data which makes sense in the light of ordinary knowledge and experience (Corbett and Firestone, 1980; Sundquist, 1978).*

Third, there are also pratical considerations which have promoted the use of qualitative methods. The increasing burden of "forms clearance" procedures required before standardized data collection instruments can be used under federal contract regulations should not be underestimated as a motivator, both for federal agencies and researchers. Since forms clearance can take from four to six months, the federal agency that asks for qualitative data in addition to quantitative can begin to "know" something about the topic in question long before a survey or testing program could begin. Thus, particularly in cases where there is only limited interest in a "bottom line" assessment, qualitative approaches may be perceived as more efficient.** Qualitative designs may also be viewed as more flexible at responding to changing policy coontexts and questions than traditional experimental designs.

It should be emphasized that the above named pressures do not provide simply a shift in emphasis from one camp to another, but rather a desire on the part of policy makers (and at least some researchers) to draw upon the best of both methods. Despite the increase in policy makers' support of qualitative research there continues to be limited interest in sponsoring true ethnographic case studies, except in the context of supplementing very large, well defined, social experiments (see, for example, Trend, 1976; Herriott, 1980). The new emphasis upon qualitative methods does not seek a paradigm shift, but would retain the strengths of quantitative research -- generalizeability of results, reliability of observations, and the ability to synthesis a large complex study in a brief report. Increasingly, there has also been a strong perceived need to address the integration of findings across different methodological

---

*That policy makers have come to view the "soft" approaches as fruitful is evidenced by the significant increase in the number of RFPs from a variety of agencies which require qualitative or case based approaches rather than (or in addition to) quantitative ones.

**That field based methods should come to be thought of as efficient is an ironic turnabout from earlier periods in which "public opinion" surveys and other survey data collection activities were touted because of their speed and and low cost.

approaches in a more formal way. Four major approaches to integrating can be identified in recent studies. Three of these -- the sequential, the parallel, and the fused models -- have been relatively widely used. One, -- the interactive model -- was developed during the course of the study of the National Institute of Education funded R&D Utilization Program, and one is relatively new.* In the remainder of the paper, the first three models will be presented briefly, and then the interactive model will be discussed in greater depth, with attention to how it evolved during the course of the R&D Utilization (RDU) study. Each of these models deals with the need to maintain some of the good characteristics of quantitative research while stressing the qualitative, or, conversely, to infuse the quantitative with some of the virtues of the qualitative. All involve "multi site, multi-method" approaches to data collection and analysis, but each represents a very distinct approach to the marriage between the two predominant methodological paradigms.

The Sequential Model

The sequential model does not alter the traditional explanations typically offered by quantitative methodologists, of the relationship between the two methodologies. Preliminary "knowing" is seen as crucial where the topic in question is poorly understood, where measurement techniques are not perfected, and where there is a need to identify or refine hypotheses. Thus, within a given study, qualitative data collection preceeds the development of survey or testing instruments, which are perceived of as the "final" data for the study. (See Zetterberg, 1962, for the classical statement of this assumption about the appropriate relationship between the qualitative and quantitative). A contemporary reflection of this approach is the "evaluability" movement, which stresses the need for a two stage evaluation of major social programs: the first involves significant field-based data collection to determine whether the treatment can be identified, as its parameters, and to develop a model of of program operations and outcomes that will form the basis for an appropriate quantitative design (see Rutman, 1980). As some have noted, this approach is particularly useful for combining formative and summative evaluation approaches: as the program develops, qualitative field data can be used to articulate issues related to design and implementation, while once it is maturely operating, a test of its achievements may be made

*I am indebted to my colleague, Dr. Catherine Batlzell, for the discussions which led to the identification of these models.

(see Sieber, Louis and Metzger, 1972). A large number of federal program studies which include an extensive "design" phase, which includes field data collection, are implicitly based upon this traditional understanding of qualitative data as "preliminary" and "hypothesis generating", while quantitative data is "final" and "hypothesis testing".

The Parallel Model

The parallel model, on the other hand, makes no such assumption about an appropriate linear relationship between qualitative and quantitative methods. Typically such designs accept the arguments of Rist (1977) or Scriven (1972) that they represent two very different "ways of knowing", and that the researcher sees and illuminates different aspects of the social phenomenon under study. Thus, for example, one would not expect an ethnographic treatment of a case to even touch upon the same patterns that the quantitative study is attempting to illuminate.

This new tradition for multi-site, multi-method case studies has typically assumed that the most appropriate approach to maximizing the contributions of both is to allow them to develop independently, but simultaneously. Thus, for example, in the study of the Rural Experimental Schools Program, 10 ethnographers resided for three years at the local sites, while centrally located "quantitative" types designed surveys, multivariate data analysis plans, and so forth (Herriott, 1980). Neither had a deep influence upon the other, because there was little intellectual communication or exchange. The quantitative and qualitative components of this project were integrated, after a fashion, but this occurred post hoc, and the materials of each group were essentially "secondary data" for the other (see Rosenblum and Louis, 1981 for examples of how the ethnographic materials were used in quantitative analysis, and Firestone, 1980 for use of the quantitative data to enrich an ethnography).

Parallelism reduced problems of coordination between competing "ways of knowing", and also eliminates many of the problems of coordination between centrally based staff and field based staff. In addition, it retains the maximum flexibility in design and analysis for both methods. It is, thus, often preferred by paradigmatic researchers, and it is frequently employed in large studies (The Rand Study of Federal Program Supporting

Educational Change, the SRI study of Teacher Corps, and the SDC study of Parent Involvement are some recent examples, in addition to the Abt Associates Rural Experimental Schools Study). However, it is not without its problems and costs. First, post hoc integration is extermely difficult, particularly where there is a great deal of qualitative and quantitative data. The process of integration is in most cases, similar to conducting a third study using secondary materials. Thus, integration across methods is usually limited, at best. Second, the parallel approach may reveal contraditions which, because of the autonomy of design, may be difficult to resolve (Trend, 1978). Finally, the approach can be very costly, since it requires that two studies, one qualitative and one quantitative, be conducted simultaneously, and resolved and integrated later.*

The Fused Model

Because of some of the limitations of the sequential and parallel approaches, there has been increasing emphasis in recent years upon an entirely different approach, which has grown up simultaneously in a number of research projects, but which is perhaps best articulated by Yin (forthcoming and 1980); McClintock, et al. (1979); and Baltzell, (1980). The new method fuses some of the most "valuable" features of quantitative data collection -- emphasis upon standardization of data points, an emphasis upon determining causality and testing hypotheses, rather than describing, and an emphasis upon cross-case analysis -- with a flexibile approach to observation and an emphasis upon holistic analysis. This approach is most frequently referred to as the standardized case method (see Baltzell, 1980). Overall, the key features of the approach are:

- considerable pre-specification of the data that is to be collected in extensive protocols, which include not only questions to be answered, but specification of documents to be collected and at least some of the respondants that must be interviewed or observed;
- built in flexibility to pursue at least some additional topics and interview different individuals from those pre-specified if these appear to be locally important to the social phenomenon under study;

*The approach being used in the ongoing SRI Teacher Corps Study, as described by Finnan (1980) may compensate, for it involved bringing field staff back into the central office for 3 weeks of each month. However, data collection and analysis apparently followed a typical parallel model.

- an emphasis upon a unit of analysis that is larger than a respondent, e.g., a social unit, a classroom, a program as it is implemented in a local context, etc.;
- an emphasis upon early data reduction and analysis while in the field, and by requiring a standardized reporting format which involves pre-identifying at least the majority of issues to be addressed;
- the use of brief, but iterative approaches to data collection. Typically there will be an initial round of field visits, a period of analysis and refinement based on these, and a second round to collect data that is missing because of changes in the design and analysis plan, or to observe changes over time; (most studies using this approach spend between one and two weeks at a given site);
- the development of causal arguments within each case and across cases using a direct replication design. The logic of this analytic approach is to find specific phenomenon in repeated cases under predictable conditions. Where the N is very small (5 or less) the approach will typically emphasize the internal validity of causal analysis within each case, where somewhat larger, the design often looks for systematic replications and attempts to establish the variability of conditions under which a phenomenon occurs.

In addition, many of the standardized case studies have taken the approach one step further toward fusion: case studies, once they are completed, are coded by the field staff, using "survey" instruments similar to those that might be administered as a respondant interview and analyzed quantitatively. (For a discussion of this as a technique, see Lucas, 1974; Yin, forthcoming; Prewitt and Eulau, 1975). Recent studies emphasizing this approach include Yin's studies of Innovations in Urban Bureaucracies (1978) and of interorganizational networks between state, local and regional education agencies (1981); King's study of staff development in desegregation; and studies of Magnet Schools as Mechanisms for Desegregation (Royster and Baltzell, 1979).

Because the standardized case method does attempt to achieve a new synthesis of quantitative and qualitative paradigms, it has generated some controversy. Ethnographers, who are socialized to believe in the appropriateness of spending long periods of time in the field and allowing all hypotheses and interpretative frameworks to emerge from field observations are appalled at the notion of standardizing observations across sites, and of using reporting formats that are pre-determined and argue that the approach cannot possibly render an accurate portrait of the social system or setting.

Quantitative researchers, on the other hand, are concerned about the development of policy recommendations from small N research, using techniques that, to say the least, do not measure up to the rigor of typical sampling and measurement that tends to be preferred. In sum, the advocates contend that the approach maximizes the value of both paradigms, while opponents contend that it bowdlerizes the intent and value of both.

Policy Researcher as Sleuth: The Interactive Model

A final new approach to integrating qualitative and quantitative data -- the interactive model -- builds upon some of the features of each of the previous three models. The major distinctive characteristics of the approach are:

- the merging of qualitative and quantitative data within as well as across sites;
- staffing patterns which involve senior researchers who participate in both quantitative and qualitative aspects of the study;
- persistant attempts to triangulate data sources and interpretations;
- cyclical interaction between the qualitative and quantitative method during all phases of the study, including sampling, instrumentation, data collection, analysis, and reporting.

Each of these features of the approach, as they were developed in the Study of the R&D Utilization Program, will be discussed briefly below.

Program and Research Context: The R&D Utilization Program was an eight million dollar demonstration effort funded by the National Institute of Education between 1976 and 1979 to promote the adoption and implementation of new curriculum and staff development materials in 300 local schools. (For more details about the program, see Louis, et al, 1979; Louis and Rosenblum, 1980; and Louis, Rosenblum, and Molitor, forthcoming). Like many demonstration efforts, there were delays in funding the research component. NIE was, however, particularly anxious not to lose data on the early development of the program, and therefore funded a regional laboratory to perform some data collection during the first months of the program, and also encouraged each of the seven demonstration projects to design relatively elaborate data

collection systems, including funding them to hire researchers to write a total of 42 case studies about particular sites, which were site specific and relatively unstandardized.

Abt Associates proposed to supplement the existing data collection activities with two waves of survey data collection, which would tap cross-site issues during early and later stages of implementation. The agency, however, expressed a preference for conducting a single survey, and asked the research staff to consider alternative ways of providing longitudinal analysis of the process of change in schools. With NIE support the research staff developed a design which, while initially considered far from ideal, finally produced a data base that allowed effective integration of qualitative and quantitative data across approximately 90 local sites involves in the program.

In the study of the R&D Utilization Program the emergenct interactive design was, in large measure, a byproduct of external features of the study context. At the time, many of these constraints were perceived as albatrosses by the research staff, who, when initiating the study, had a marked preference for the sequential model described above. However, in the process of coping with external constraints, problems with availability of data, and data quality, a great deal was learned about how to maximize the utility of different approaches, "found" data, and both ad hoc and systematic information. In other words, the design that emerged is robust against many of the "normal crises" which occur in field based policy research. The specific elements of the design will be discussed below.

Key Features of the Methodological Approach: First, the approach used in the study utilized all three of the models discussed above. The sequential approach was utilized, in that the single survey that was conducted in the portion of the study concerned with school change processes and outcomes was developed based on the qualitative data collection that preceeded it. The parallel model was used also, in that a standardized data collection operation completed by AAI staff, occurred simultaneously with the development of 42 "mini ethnographies" (Knapp, 1977) written by a field staff who had been previously hired by the funded projects and who were essentially independent of the emerging perspectives of the central research staff. Finally the

qualitative field data collection by AAI used the standardized case model for an additional 52 sites including the quantification of case materials.

The integrated approach, however, is more than simply utilizing a variety of models for integrating qualitative and quantitative data across sites. Our approach introduced several new techniques for integration that are distinct from those discussed above. These included:

(1) Merging Qualitative Data Sources Within Sites: The Consolidated Coding Process: Much of the discussion of integrating qualitative and quantitative methods involves essentially cross-site analytic issues. The challenge facing us included the cross-site merger of different data sources, but even more pressing was the problem of a diversity of within site data sources. One feature of policy and applied research is that it is frequently very "messy". For example, most major programs are not studied simply by one group of "evaluators" or researchers. Rather, there may be internal evaluations, reporting and administrative data collection from the funding agency, and a sequence of external researchers. The multiplicity of research and reporting requirements is a source of major concern to those who are being studied, who may frequently plaintively complain that "someone was here just last week asking me the same question". This was a particular problem in the RDU program, where "mini ethnographers", NIE administrative reporting requirements, two sets of external evaluators, and a variety of otner researchers interested in the program all descended with regularlity upon the schools involved, but is increasingly a dilemma in other programs, such as Title I, IVC and others.

For any site, our data could include any or all of the following:*

- demographic data;
- survey data of "key informants" collected at the beginning of the program by the first evaluation contractor;
- "event triggered" reports, discussing different phases of the school's progress through the project, which were management reports to NIE for monitoring purposes;

*Two other data sources were available at a later point in the study:

- a survey of the principal;
- a survey of between five and ten teachers who were targeted users of the innovation adopted by the school.

- our own field reports; which included a standardized case report and a quantified survey report filled out by the site visit team;
- a "mini ethnography", and a "case study writers survey" which attempted to parallel some of the coding categories that were used by the AAI field staff;
- various documents, provided either by the project, the local site, or other persons.

In no more than 20 percent of the sites was a complete data base available, and in most cases major instruments or documents were unavailable. In sum, we had a "missing data" problem of the first magnitude. Yet, it seemed foolish to throw away "evidence" of any sort, particularly given the high level of effort that had gone into collecting some of the data that was missing for the largest number of sites.

Rather than analyzing each data source separately, it was decided to combine all data sources for each site in the "intensive" sample, e.g., a site where we had a "mini ethnography", or which had been site visited by an AAI staff member. This was done by developing a coding form, in which a single score or rating was given for each variable on the basis of a senior researcher's judgment derived from all the possible data sources enumerated above (with the exception of the formal principal and teacher surveys. The level of measurement was geared to maximizing the completeness of available data. Thus, for example, the standardized case materials stressed precise measures of level of effort devoted to planning and implementation, in terms of staff days. The case studies tended to have much less precise measures. Level of effort was, therefore coded as "high-medium-low", where these were defined as ranges of staff days. Responding to these site based questionnaires which included 240 dichotomous or Likert scale items was an extremely time-consuming task, involving between two and three days of reading materials, and verifying responses on the coding form. All coding was conducted by core senior staff members, who had made visits to at least four of the sites, and who went through an intensive two day session in which common interpretations of items were reached.

While the process of "sleuthing" through the data trail for each case did consume a great deal of time, what resulted was a quantitative data

base for 80 sites, which covered issues that could not be easily tapped through traditional survey methods, such as the "quality" of the decision-making process, patterns of influence of different actors over decisions, and so forth. Because of the diversity and large number of sites, we do not believe that any other form of data reduction and analysis would have produced a data base which at once reflected the "holistic" knowledge that site visit teams brought to the cases, but also the reliability of standardized data, integrated both within and across sites.* And, while time-consuming, it was considerably less costly than other alternatives, such as (1) ignoring non-conforming data sources; or (2) gathering additional data on site in order to develop more complete standardized case studies.

(2) Staffing: The two cultures of research often result in research designs where staff members are specialized, either in qualitative data collection and analysis, or quantitative data collection and analysis. This approach, while appropriate for the parallel model, does not facilitate integration and interaction. The staffing patterns on the RDU study involved having the same core staff of senior researchers involved in all instrumentation, data collection and analysis. Thus, every individual who contributed as a major author or analyst to the study was personally familiar with site processes at all levels, and had responsibility for some portion of the quantitative analysis in the study. This "integration by staffing" is, perhaps, one of the most effective ways of ensuring that the cycle of testing quantitative and qualitative observations occurs on a regular basis.

(3) Triangulation and Data Quality: As Webb, et al (1963) have pointed out, qualitative data becomes more compelling if observations are supported by multiple sources of evidence, or observations. The issue of reliability was of deep concern to us, in part because of the rather motley nature of our underlying data sources. Thus, our approach involved several approaches to triangulation:

- inter-observer: site visits were conducted in teams. Teams were required to reach consensus in preliminary rating of sites on quantitative dimensions. In addition, when using the consolidated coding form, inter-rater reliabilities were conducted;

*A limited inter-rater reliability check was conducted, which revealed an agreement rate of 72 percent. Several sites were eliminated because of the amount of missing data.

- holistic vs. categorical: as part of the interweaving of qualitative and quantitative data, findings from the qualitative data base were constantly tested out in the quantitative and vice versa. For example, in attempting to develop a categorical variable summarizing "site success", a definition of success was first discussed among the staff on the basis of the field data visits. An indicator that reflected these discussions was developed from the survey data of principals and teachers. The "intensive" sites were then categorized, using the quantitative indicator, and staff again discussed whether they were correctly classified, using the "holistic" judgements that they developed in the field visits, or from reading the "mini ethnographies".

- case vs. survey: in several cases, similar measures were built into teacher and principal surveys, and into the consolidated case coding instruments. Correlations between the perceptions of local respondents, and the perceptions of case raters were calculated, and several variables were discarded where correlations were not positive and significant.*

- survey vs. survey: in all cases we were trying to obtain building level measures of the process and outcomes of implementing new curriculum and staff development materials. This meant aggregating teacher responses to the building level. In order to ensure that we were not falling into the "ecological fallacy", an analysis of variance between and within buildings, using both teacher and principal data, was performed.

  After the latter two activities were completed, robust measures were scaled to form a single school level score, reflecting the responses of principals, teachers and the external case material collected by field staff.

(4) Cyclical Interaction Between Quantitative and Qualitative Data: In the sequential and standardized case models, the assumption is that qualitative data will influence quantitative data collection and analysis, but not vice versa. Most discussions of integrating quantitative and qualitative methods also promote integration at one phase of the execution of the study. Thus, for example, the sequential approach promotes integration during the period of instrumentation and final design for the "final" quantitative data collection; the parallel approach only after all data collection

*For a further discussion of case-survey triangulation, see Louis, Rosenblum, and Molitor, forthcoming; Appendix B.

and analysis in the separate streams is completed, and the standardized case approach during the process of cross-case analysis, during which counts and frequencies (or even correlations) are made.

In the RDU study, on the other hand, quantified data was available from a very early point in the study, both as a result of the surveys that were conducted prior to the award of the contract to AAI, and also because the first round of field visits were coded while we were on site. The results of early descriptive analysis changed the focus of our qualitative data collection in significant ways, for they pointed out topics where there was little variation between sites (and certain issues that were expected to dominate the case studies were therefore eliminated). In addition, they suggested some issues to look for -- such as the relationship of racial diversity to the process of change -- that were not initially picked up on by the field visit or case study writers as significant. This cyclical interaction between qualitative and quantitative occurred throughout the study, during staff analysis seminars and during the construction of instruments at various phases of the study. Cyclical integration stresses the following features:

- interaction between qualitative and quantitative in sampling: the interactive approach allows purposive sampling for cases to be combined with random sampling (or sampling of the universe) for survey or other structured data collection. This feature maximizes discovery and generalizeability.
- interaction between qualitative and quantitative in instrumentation: a constant interaction between qualitative and quantitative analysis procedures produces an iterative approach to instrumentation, both for field data collection and more standardized coding instruments or data collection. Iterative instrumentation would be almost impossible for most standard longitudinal research designs, which involve repeated measures. This problem is avoided in the "consolidated coding" procedure, in which repeated measures are reconstructured from a broad evidentiary base.
- interaction between qualitative and quantitative in data collection: the development of survey-type instruments for use even in early field data collection forced the research team to seriously think through measurement assumptions at a stage where, in a more typical sequential or parallel study, it would not occur. We believe that this contributed significantly to the quality of our measurement.

- interaction between qualitative and quantitative in analysis: in the interative model, it is impossible to identify a distinct "analysis phase" in the research project. Rather, analysis begins with the first data collection event, and occurs periodically throughout the project. More importantly, however, is the fact that the simultaneous analysis of quantitative and qualitative data by the same staff continuously require a testing and verification of one data source against another, increasing both reliability and validity.
- interaction between qualitative and quantitative in reporting: while some of our reports are more "qualitative" and others more "quantitative" the immersion of all staff in both kinds of data have meant that no report draws solely upon one data source.

Some Reflections on the Potential Strengths and Weaknesses of the Interactive Model for Social Policy Research

Since there has only been one "test" of the approach that we have described above in a major educational study, any assessments of its viability must be preliminary. However, it is appropriate to draw some conclusions about what has been learned regarding the integration of quantitative and qualitative methods in design, data collection and analysis.

Design: Since the beginning of the "evaluation" movement, many disciplinary researchers have complained about the lack of attention on the part of policy makers to designing "evaluable" programs, and their tendency to involve researchers too late, and at too low a level of effort to carry out the optimal research designs (Rodman and Kolodny, 1964, Weiss and Rein, 1969). Unfortunately, most of these please have not been attended to, for a variety of reasons which are often beyond the control of the research branches of various government agencies. Thus, as policy researchers, we are typically forced to choose between a research context that is less than ideal, or not conducting research on interesting and significant policy endeavors. The "sleuthing" strategy that characterizes the interactive model is premised on the assumption that the least promising circumstances can yield useable and even exciting data, that almost any piece of information can be turned into a "clue" to understanding the phenomenon in question, and that systematic analysis of clues is important.

The interactive approach should not necessarily be viewed as the ideal research design, for it lacks the elegance or simplicity of a true paradigm. However, it is particularly well suited to addressing some of the realities of policy research which often causes stresses and strains in more elegant designs. Among these conditions are:

- where the study combines both significant exploratory and/or evaluative and hypothesis testing components;
- where it seems important to have both a "rich" or holistic understanding of a process occuring in a field setting, and a broader cross-case analysis which addresses some of the same issues;
- where the policy audience is composed both of people who prefer qualitative "valid" data, and quantitative "generalizeable and reliable" data;
- where the field reality is "messy": lots of previously collected data exist but there is a great deal of variance in quality and depth of information between sites, but where response burden is a significant practical or political issue; where resources may be too limited to begin afresh with a totally new design; or where the research involves documenting a longitudinal process that is already well underway.

The interactive approach is not a simple one however, and causes its own problems. It requires constant attention from staff members, committed, consistent, in-house senior staff, each of whom is capable of both qualitative and quantitative data collection and analysis. One feature of the approach that should be emphasized is the need to maintain very low rates of turnover among project staff: since the process of design, data collection and analysis are intertwined, any need to replace a key person involves considerable costs in socialization. In fact, our attempts to replace key staff members during the last year of the project were not at all successful; and it proved necessary to extend our work over a longer period of time in order to complete the analyses. In addition, it must be emphasized that staff members themselves must be relatively free of paradigmatic preferences. If either the project director or other key staff members are reluctant to become equally involved in qualitative and quantitative analysis, promoting cyclical integration may be extremely difficult.

Data Collection: The wide variety of different types of data collection activities that were carried out in the RDU study allows us to address

issues regarding validity/reliability that are often debated among those who are concerned with the increase in multi-site qualitatively based studies.

One of the major informal debates occurring among qualitative cross-site researchers is the question of how one can best preserve the "holistic" insights obtained in a traditional ethnography with the necessary truncating of time in the field which occurs as financial resources are being stretched to cover many sites. The proponents of the "standardized case study" tend to prefer brief on-site data collection activities which are conducted exclusively by in-house staff members (e.g., a staffing approach that is more similar to traditional quantitative data collection procedures than to the true ethnographic approach). The individuals who argue for "mini ethnographies", on the other hand, contend that even a superficial understanding of the functioning of a program, organization, or other social unit in a local setting must involve more time on site, more visits over a period of time, in addition to greater flexibility in observations. Typically a "mini ethnography" approach argues for hiring trained individuals -- often consultants -- who are closer to the system in question, so that they can use whatever time is available for observation and interviewing in a way which is more responsive to the unfolding of events at that site.

Our study provided us with the opportunity to assess the "richness" and depth of data (validity) collected using a "mini ethnography" approach and the standardized case approach.

The typical mini ethnography in our study consumed about 15-20 days of data collection and very preliminary within-site analysis on the part of the case study writer, who was, on average, a university professor who took on the job of writing the case studies out of personal interest.* As a result, there was, in a number of cases, a gread deal of "contributed" time. The typical standardized case study, on the other hand, consumed approximately 4-5 person days of data collection, which occurred at a single point in

*The appropriateness of the background of case study writers for conducting this type of research varied considerably. In addition, some of the 42 "mini ethnographies" actually used a standardized case approach.

time.* Two person teams were used, and the team conducted some joint interviews with pre-identified key informants, and some individual interviews with teachers and other relevant respondents. Site cases were written by one team member on the basis of a team debriefing and discussion; each case was reviewed by the other team member and augmented where necessary.

On the whole, we believe that our standardized case studies yielded data that was equally insightful into local site specific processes as the more intensive mini-ethnographies. In addition, it was typically far more useful for the purposes of cross-site analysis. In a number of instances, we obtained both standardized cases and "mini ethnographies" on the same site. Where there were differences in what was emphasized in each case, in most instances our data were equally detailed, and equally informative about the longitudinal processes of change. In no case where we were familiar with the mini-ethnography sites, did we find any major contraditctions between the data that we had collected, and that collected and reported through more intensive, less standardized means.

Two conclusions may be drawn from this informal comparison, which was corroborated not only by our possibly biased project staff, but also by several individuals who served as consultants to the study who were familiar with both types of quantitative case data. First, the more costly "mini ethnography" does not result in sufficiently better data to justify the additional use of resources. Second, the standardized case study can reveal local site uniqueness (if the reporting format allows this) and, in addition, is far more useful for cross-site synthesis. Mini-ethnographies are no better at reflecting "holistic" patterns within the site because

---

*It should be emphasized that the 4-5 days per site that yielded relatively valid data for our study should not be generalized to other settings. For example, when examining district-wide behavior, it is clearly necessary to have a somewhat larger number of person days than when examining program functioning in a single school. However, the general principal of contrasting mini-ethnographies and standardized case materials still holds.

Further, it should also be emphasized that information gathered during the 4-5 days on site was augmented by review of documents prior to and after the visit, and through familiarity with the nature and substance of the intervention gained in interviews with managers at the funded project level and with the external providers of technical assistance.

they still lack the essential feature of a true ethnography: a continuous, on-site presence which involves substantial participants observation in addition to interviewing.

Validity is not the only problem, however. Those who are concerned about the increasing use of qualitative cross-site analysis often point to the issue of reliability of field or ethnographic observations as a serious one. This concern may be even greater when one is attempting to address very slippery and non-tangible features of organizational process such as "the influence of the principal over the decision-making process". We believe that the deliberate attempts within the study to compare the data obtained from qualitative methods with more traditional survey methods indicate that, with care, this should not be regarded as quite as serious a problem as has often been thought. With adequate triangulation, and a stable staff that shares and is fully socialized in common definitions of the phenomena under study, it is possible to semistructure structure data collection techniques and still achieve high levels of agreement between different sources of data.

Let us take, for example, one of our less concrete measures of school level outcomes, "overall organizational impacts on the school". After the first round of preliminary site visits, it was decided by the staff that there were many spin-off effects in organization development and improvement that were not directly tied, in all cases, to the implementation of a new curriculum package. These included changes in staff morale, improvement in the image of the school in the community, and the like. We then developed a set of Likert scales reflecting side benefits that we had observed, and during our major qualitative data collection we rated all of the schools that were visited on this battery. The battery was also included in the surveys of teachers and principals. While the field staff tended to rate the schools somewhat less generously than indigenous respondants on broad school organizational impacts, the ratings that were given by the field staff correlated .55 with the principal ratings, and .44 with the teachers. These are quite typical of the intercorrelations between data sources that were achieved for items or indices which overlapped.

We were, finally able to achieve even greater reliability in our quantitative data set by combining school scores on our outcome variables from both originally qualitative sources (e.g., from the "consolidated coding form") and from survey data.*

Integrating Qualitative and Quantitative in Analysis: As has been implied repeatedly above, one of the main features of the interactive approach is that this analysis does not occur after data collection is implemented. Rather, analysis is an incremental feature associated with each site visit, between groups of field data collection activities, before the design of later instruments, and so forth. It was this constant process that permitted the research staff, as a group, to arrive as the consensus necessary to develop reliable, holistic cross-site coding schemes, which were the key feature around which the integration of qualitative and quantitative occurred.

The process of trying to quantify our field data at the time during which it was being collected was a key feature of the integration process. Each time a field site team was faced with the necessity of making judgements about the local site processes on a Likert scale, it required them to clarify their perceptions about the measurement properties of the scale, the phenomena that had been observed, the relative weight to put upon different respondent's perspectives on the issue in questions, and whether the most important features of what was happening at the site were not being tapped by the site coding instruments as they existed. This process, which occurred both in the field and during extensive debriefing sessions with the entire core research staff of eight ensured that (1) we were constantly pressing to have our quantitative instruments reflect, as much as possible, our more holistic understanding of how sites were operating; (2) we were able to develop through quantification, a more precise understanding of the site that was reflected in the level of detail in each case write up; (3) the existance of the quantitative forms continuously pressed us toward the standardization that was necessary in order to adequately prepare for a cross-site analysis of a very large number of sites.

*For a more extensive discussion of the "quality" of the quantitative data that was derived from various sources, see Appendix B, in Louis, Rosenblum and Molitor, forthcoming.

In the end, our analysis cannot be said to be either quantitative or qualitative. For example, can a data based composed of numbers that is entirely dependent on the iterative, holistic judgements of experienced site field teams be described as only quantitative? While the analysis procedures used to manipulate the data are statistical, the data itself, and any interpretation of results, is totally conditioned by its origins. On the other hand, as we approach any given analysis using case materials rather than quantified data, it has become genuinely impossible not to embed that activity in our knowledge of the descriptive statistics and correlational relationships that were available to us well before qualitative data collection had ended.

Based on the previous experience of the RDU study and their reports, many colleagues who have faced the problem of cross-site analysis with N's of ten or more, it is impossible to avoid quantification of the qualitative data. The process of holistic analysis appears to break down at some point, and the analyst begins jotting down counts of occurances of phenomena, possibly even computing rank order correlations, but unquestionably thinking as a quantitative analyst. At some point -- perhaps where the N reaches about 15 -- the ability of most people to hold the holistic story together across cases diminishes. There are two responses in analysis. The most typical one, perhaps, is to essentially throw away all but the best or most familiar data informants description. At this juncture, some formal quantification of key variables may help. When, as is increasing the case in policy research, the number of sites is 25 or more, and the number of informants numbers in the hundreds, we are unquestionably in a condition where formalization of data analysis procedures -- e.g., some form or quantification, whether highly rigorous or less so -- is essential to simply manage the data.

Based on the experiences of the RDU project, it is highly recommended that such quantification occur <u>early</u> and <u>iteratively</u> in order to most effectively integrate the process of more holistic site specific analysis and the cross-case analysis which is necessarily more attentive to variables and cross sections of a larger process and structure. More importantly is the fact that the simultaneous analysis of quantitative and qualitative data by the same staff continuously require a testing and verification of one

data source against another, increasing both reliability and validity. The alternative is throwing away the data that cannot easily be incorporated -- a procedure that may add to the theoretical compact elegance of the report, but which does little to further the development of cross-site analysis as a legitimate methdology.